

R U 2 1 3 8 0 7 6 C 1



(19) RU (11) 2 138 076 (13) C1

(51) МПК⁶ G 06 F 17/30

РОССИЙСКОЕ АГЕНТСТВО
ПО ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ РОССИЙСКОЙ ФЕДЕРАЦИИ

(21), (22) Заявка 98118879/09, 14.09.1998

(24) Дата начала действия патента 14.09.1998

(46) Дата публикации 20.09.1999

(56) Ссылки: WO 98/20434, 14.05.98, US 5748954 A, 05.06.98, US 5578954 A, 19.11.96.
Управляющие вычислительные машины в АСУ
технологическими процессами/Под
ред. Т.Харрисона. - М.: Мир, 1975, с.48 - 52,
фиг.1.11.

(96) Адрес для переписки
115446, Москва, Коломенский пр-д 1а, ЗАО
"МедиоЛингва", Ашманов И.С.

(54) СИСТЕМА ПОИСКА ИНФОРМАЦИИ В КОМПЬЮТЕРНОЙ СЕТИ

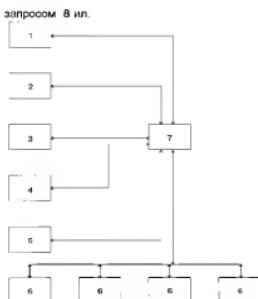
(57) Рефэрят:

Изобретение относится к вычислительной
технике. Технический результат заключается
в расширении функциональных возможностей
и повышении эффективности использования
оборудования, контролирующего и обрабатывающего
запросы в компьютерной сети. Для этого
также что система поиска информации в
компьютерной сети содержит устройство
сбора и обработки документов и память
документов и адресов. Система содержит
также устройства управления абонентов
и индексную память. Устройство поиска
включено в сеть и включает блоки хранения
помощных индексных образов и запросов на
поиск из устройств управления абонентов,
преобразования запросов на поиск в
индексный образ, сопряжения с индексными
образами индексной памяти, выбора адресов
документов, соответствующих индексному
образу абонентского запроса, ранжирования
выбранных документов и формирования
индексному образу абонентского запроса и
формирования списка найденных документов
с указанием их адресов в компьютерной сети
в соответствии с указанным абонентским

(71) Заявитель:
Закрытое акционерное общество
"МедиоЛингва"

(72) Изобретатель: Серебренников О.А.,
Ашманов И.С., Иванов А.П., Пашко Д.А.

(73) Патентообладатель:
Закрытое акционерное общество
"МедиоЛингва"



R U 2 1 3 8 0 7 6 C 1

Фиг. 1



(19) **RI** (11) 3138076 (13) C1

(51) Int. Cl. 6

RUSSIAN AGENCY
FOR PATENTS AND TRADEMARKS

(12) ABSTRACT OF INVENTION

(21), (22) Application 98116879/09, 14.09.1998

(24) Effective date for property rights 14.09.1998

(46) Date of publication 20.09.1999

(98) Mail address:
115446, Moskva, Kolomenskij pr-d 1a, ZAO
"MediaLingva", Ashmanova I.S.

(71) Applicant:

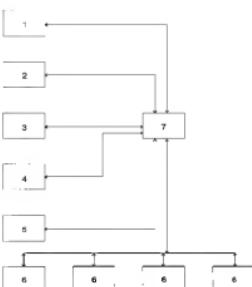
(72) Inventor: Serebrennikov O.A.,

(73) Proprietor:
Zakrytoe aktsionernoe obshcheschestvo "Medial inc."

(54) DATA RETRIEVAL SYSTEM IN COMPUTER NETWORK

(57) Abstract

FIELD: computer engineering **SUBSTANCE:** system has document acquisition and processing device as well as document and address memory. It also has subscriber control devices, document indexing devices, and index memory. Data retrieval devices are used for receiving search requests and search requests coming from subscriber control devices, converting search requests to index image, comparing it with index images of index memory, retrieving addresses of documents corresponding to index image of subscriber request, ranking chosen addresses according to their correspondence to index image of subscriber request, and forming list of found documents including their addresses in computer network according to mentioned subscriber request. **EFFECT:** enlarged functional capabilities and improved efficiency of hardware use. 8 ovng



RU 2 1 3 8 0 7 6 C1

R U 2 1 3 8 0 7 6 C 1

Изобретение относится к области вычислительной техники и может быть использована для поиска информации в компьютерной сети, например, Интернет, по абонентскому запросу на естественном и/или формальном языке.

Известны устройство и способ поиска релевантных документов (см. патент США N 5575954, опубл. 19.11.96 г.) осуществляющие сравнение различных документов и ранжированием их по степени соответствия. Известная система предусматривает обозначение в сеть только на формальном языке.

Известен способ и устройство поиска документов в памяти, их классификации и размещения в определенном порядке (см. патент США N 5748554, опубл. 05.05.98 г.).

Известная система предусматривает обозначение в компьютерную сеть только на формальном языке.

Известная система для связи абонентов в Интернете, сбора и отображения информации (см. заявка РСТ ВО 98202434, опубл. 14.05.98 г.) Известная система выбрана как наиболее близкий аналог к патентуемой.

Известная система позволяет обращаться в компьютерную сеть только на формальном языке.

Задачей патентуемого изобретения является обеспечение возможности обращения в компьютерную сеть с запросом на поиск документов любым образом, в т.ч. на естественном языке.

Достижимый результат состоит в расширении функциональных возможностей и повышении эффективности использования оборудования системы поиска информации в компьютерной сети.

Результат достигается тем, что в систему поиска информации в компьютерной сети, содержащую устройство сбора и обработки документов, с возможностью приема документов из памяти, содержащей в полученных документах адресов, приема документов в соответствии с выделенными адресами, обмена документами и ссылками в памяти документов и адресов, анализа наличия документов и адресов в памяти документов и адресов, изменения и удаления документов из памяти документов и адресов по результатам приема адресов, поиска документов в памяти, содержащей введенную устройством индексированную информацию о введенных и удаленных адресах, приема и выдачи документов и адресов устройства управления абонентов с возможностью формирования запросов на поиск на формальном и/или естественном языке, приема и удаления документов и адресов, заголовков, атрибутов, аннотаций документов, а также устройства, содержащего устройство введенной устройством индексации документов с возможностью формирования индексов документов, выделение заголовков и аннотаций документов и их частей, выделение атрибутов документов, формирования индексного формата документа, формирования индексного образа документов, формирования индексного формата, адресов документов и атрибутов документов, приема адресов измененных или удаленных документов, удаления индексных образов документов, передачи поступивших адресов и сформированных индексных образов в индексную память индексную память с возможностью приема адресов и

индексных образов, передачи индексных образов в устройство поиска, устройство поиска, возможное выполнение приема индексных образов и запросов на поиск от устройства управления абонентов, преобразования запросов на поиск в индексный образ, сравнения с индексными образами индексной памяти, выбора адресов документов, соответствующих индексному образу абонентского запроса, ранжирования выбранных адресов по степени соответствия идентификаторов документов, и формирования списка найденных документов с указанием их адресов в компьютерной сети в соответствии с указанным абонентским запросом.

5 Конкретный, но не ограничивающий патентуемое изобретение, пример выполнения системы приведен на фиг. 1-8, где указано устройство 1 сбора и обработки документов, устройство 2, содержащее и идентификаторы, устройство 3 индексирования документов, индексная память 4, устройство 5 поиска, устройство 6 управления абонентов, устройство 7 механизма обмена, устройство 1 содержит блок 8 микропроцессорного управления, буферную память 9 документов, блок 10, в котором

20 адресов документов, таймер 11, блок 12 буферной памяти 13 документов, коммутатор 14, блок 15 микропроцессорного управления, содержит память 15, узел 16 управления и обработки, узел 17 ввода-вывода памяти 2 содержит блок 18 микропроцессорного управления, блок 19 памяти документов, блок 20 памяти адресов, таймер 21, коммутатор 22, блок 3 индексирования содержит блок 23

25 микропроцессорного управления, блок 24 выделения заголовков документов, блок 25 выделения аннотаций документов, блок 27 выделения атрибутов документов, блок 28 формирования индексов документов, блок 29, буферную память 30 адресов, таймер 42, буферную память 30 адресов, блок 24

30 индексирования документов, коммутатор 32, индексная память 4 содержит блок 33 микропроцессорного управления, блок 34 памяти индексов, блок 35 памяти атрибутов, блок 36 памяти адресов, коммутатор 37, устройство 38, устройство 39, устройство 40, блок 45

35 микропроцессорного управления, блок 39 фиксации запроса, блок 40 фиксации атрибутов, буферную память 41 индексов, блок 42 сравнения, память 43 адресов, блок 44 ранжирования, блок 45 формирования списка устройств 6 управления абонента содержит блок 46 микропрограммного управления, блок 47 устройства и обработки, блок 48 формирования запроса, блок 49 фиксации ответа, блок 50 анализа.

40 Все узлы, входящие в состав системы, могут быть реализованы с помощью программно-аппаратных средств, основанных на микросхемах, например, фирмы Intel. Система работает следующим образом. Устройство 1 собирает и отбирает из первичных документов, производит копирование документов, и их адресов в память 2, контролирует изменения документов и осуществляет удаление документов и адресов из памяти 2, если это необходимо.

45 Работа устройства осуществляется под

RU 2 1 3 8 0 7 6 C1

управлением блока 8, в котором функции контроля, анализа и обработки осуществляют узел 18 в программе 17, блок 15, обмен информацией осуществляется через узел 17.

Документ из выбранного сервера считывается через узел 17 в памяти 9, из которой блоком 10 выбираются адреса связанных документов и передаются в памяти 13. Адреса из памяти 13 и памяти 2 сравниваются в блоке 8 и, если обнаруживается совпадение, что указывает на наличие документа в памяти 2, то передаваемая информация из памяти 9 через коммутатор 14 в памяти 2 не происходит и в следующем цикле работы памяти 9 эта информация просто стирается. Если адрес не обнаружен в памяти 2, то осуществляется передача документа из памяти 9 через коммутатор 14, узел 17 и устройство 7 в памяти 2. Устройство 1 управляет серверами коммутатора - если обнаруживает адрес документа, подлежащего удалению или изменению, то выбирает его из памяти 2 и производит соответствующую процедуру удаления или замены.

Процедура замены или удаления документа может быть также инициизована таймером 3. По истечении времени таймера производится выборка адреса документа из памяти 2, который через коммутатор 14 передается в узел 17 ввода-вывода. Далее производится проверка наличия документа по заданному адресу. Если оказывается, что документ был обновлен, производится замена новой версии документа, а предыдущую сохраняется в памяти 2. Если документ по заданному адресу отсутствует, информация о нем удаляется из памяти 2.

Аналогичную процедуру осуществляет устройство 1, если при повременном сканировании таймером 11 адресов, выбираемых из памяти 13, обнаруживает адрес, указанный на изменение или удаление.

Память 2 работает под управлением блока 18, имеющего структуру аналогичную блоку 8. Принимаемые документы и адреса хранятся в блоке соответственно 19, 20 и, при необходимости, передаются через коммутатор 22 и узел ввода-вывода блока 18 в устройство 6, а также по требованию Таймера 3, передаются в память 2 адрес и соответствующий документ, который необходимо удалить или изменить.

Любой поступающий в систему документ подвергается некоторым преобразованиям для формирования индексного образа. Это осуществляется в устройстве 3, работающем под управлением блока 23, структура которого изображена на рисунке 6. Структура блока 23 показывает из памяти 2 выборку новых адресов и документов. При этом исходные тексты документов передаются на входы блоков 24, 25, адреса документов - в блок 30, а адреса удаленных документов - в блок 30. Из документов выделяются заголовки блоком 31. Из документов, блоком 26 формируется индексный фрагмент блоком 26 путем статистической обработки документов и атрибуты документов блоком 27. Эти компоненты в блоке 28 объединяются в поисковый индексный образ и совместно со ссылкой документа из памяти 30 передаются через коммутатор 32, узел ввода-вывода

блока 23 и устройство 7 в память 4. Если обнаруживаются ссылки удаляемых или изменяемых документов, то они поступают в память 31 и далее в память 4, где осуществляется выборка соответствующих индексов и их последующая обработка. Индексы могут быть подвергнуты обработке и по результатам временного сканирования с помощью таймера 29.

В памяти 4 хранятся индексные образы документов из блоке 34, атрибуты - в блоке 35, адреса - в блоке 36, а также в блоке 36, информация из которых может передаваться через коммутатор 37. Работа памяти 4 осуществляется под контролем и управлением блока 33, имеющего структуру аналогичную структуре блока 8.

Работа устройства 5 осуществляется под управлением блока 38, имеющего структуру, аналогичную структуре блока 8. Через узел 39, а в блок 40 и память 41 поступают из памяти 4 соответственно атрибуты документов и индексные образы документов. В блоке 42 осуществляется сравнение записей абонента и, в случае необходимости, производится обработка с помощью индексами документов. Результаты сравнения фиксируются в памяти 43 и синхронизируются в блоке 44 с использованием информации от памяти 41. Результаты ранжирования фиксируются блоком 45 и передаются в устройство 6 соответствующим абонентом.

В устройстве 6 работает управляемый и контролируемый блок 46, структура которого идентична структуре блока 8. Кроме того, осуществляется ручное управление и контроль с помощью пульта 47. Запрос абонента формируется в блоке 48, ответ из устройства 5 фиксируется в блоке 49 и с помощью блока 50 осуществляется анализ ответа и, если требуется, формируется дополнительный запрос в блоке 51.

Таким образом, система позволяет проводить обзор и обнаружение информации о документах, хранящихся в компьютерной сети, сохранять информацию о документах в специализированной базе данных индексного хранилища, производить поиск документов в компьютерной сети на основе соответственного запроса, формировать итоговую информацию о документах, а также по набору различных атрибутов документов, ранжировать список найденных документов по степени соответствия запроса пользователю, получать информацию о документах сети без непосредственного обращения к ним.

Это обеспечивает расширение функциональных возможностей и повышение эффективности использования оборудования системами поиска информации в компьютерной сети.

Формула изобретения:

Система поиска информации в компьютерной сети, содержащая устройство обработки и обработки документов с возможностью приема документов из серверов сети, выделения в полученных документах адресов приема документов в соответствии с выделенными адресами, обмена документами и ссылками с памятью документов и адресов, анализа наличия документов и адресов в памяти документов и адресов, изменения и удаления документов

R U ? 1 3 8 0 7 6 C 1

из памяти документов и адресов по результатам анализа адресов, память документов и адресов, изменения памяти документов и адресов, изменения документов и памяти документов и адресов по результатам анализа адресов, память документов и адресов с возможностью приема и выдачи документов и адресов, устройства управления абонентов с возможностью формирования запросов на поиск на формальном или естественном языке, приема и удаления документов и адресов, запросов на атрибуты документов, отличаясь тем, что в нее введен *устройство индексирования документов с возможностью формирования индексов документов, выделения заголовков и аннотаций документов и их частей, выделения атрибутов документов, формирования индексного формата документов, формирования индексного образа документов путем копированием индексного формата, адресов документов и атрибутов*

документов, приема адресов изменяемых или удаляемых документов, удаления индексных образов документов, передачи поисковых запросов и сформированных индексных образов в индексную память, индексная память с возможностью приема адресов и индексных образов, передачи индексных образов в устройство поиска, устройство поиска с возможностью приема поисковых индексных образов и запросов на поиск от устройства управления абонентов, приема и удаления документов и адресов в индексный образ, сравнения с индексными образами индексной памяти, выбора адресов документов, соответствующих индексному образу абонентского запроса, ранжирования выбранных адресов по степени соответствия индексному образу абонентского запроса и формирования списка найденных документов с указанием их адресов в компьютерной сети в соответствии с указанным абонентским запросом.

5

10

20

25

30

35

40

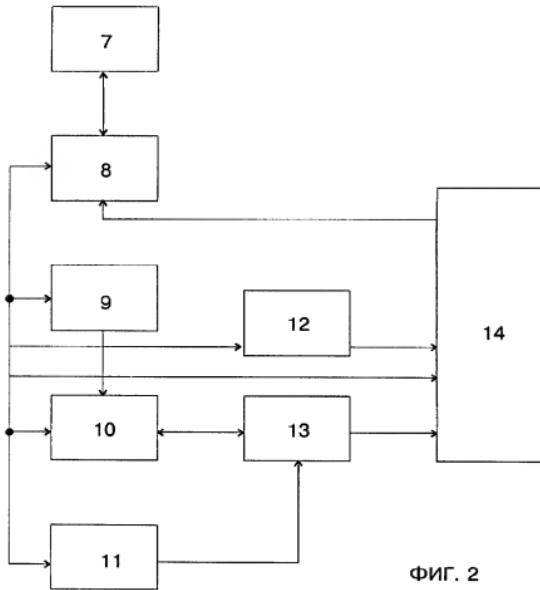
45

50

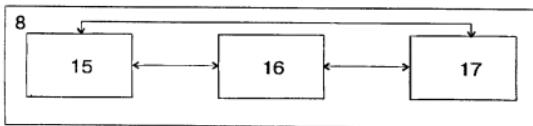
55

60

R U 2 1 3 8 0 7 6 C 1

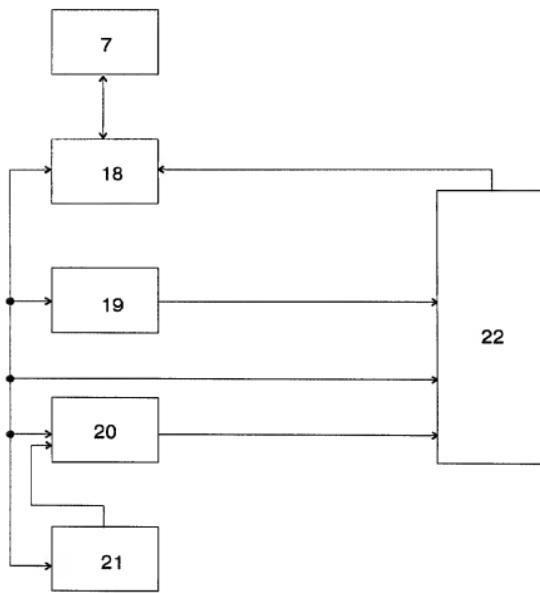


ФИГ. 2



ФИГ. 3

RU 2138076 C1

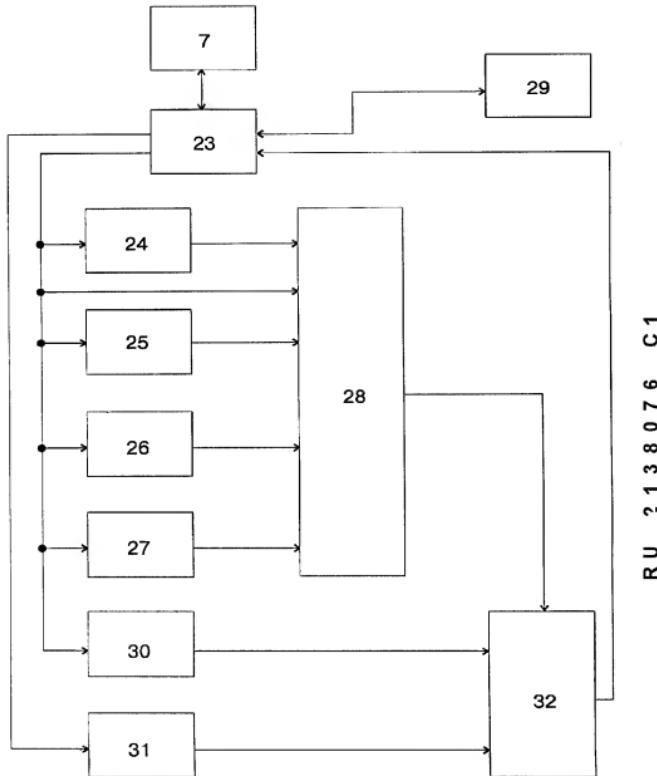


Фиг.4

РУ 2138076 С1

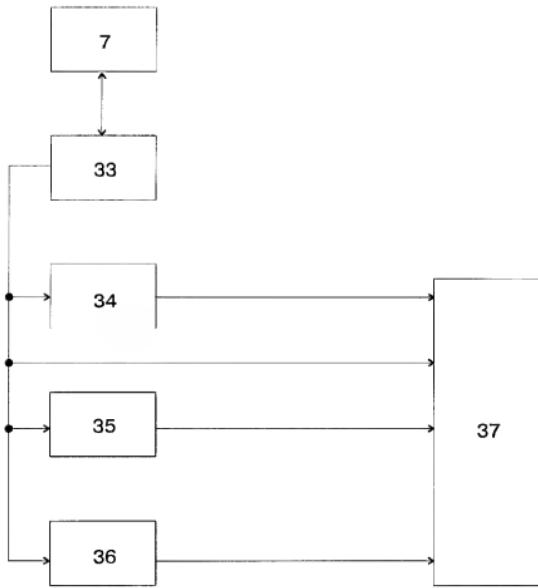
РУ 2138076 С1

РУ 2138076 С1



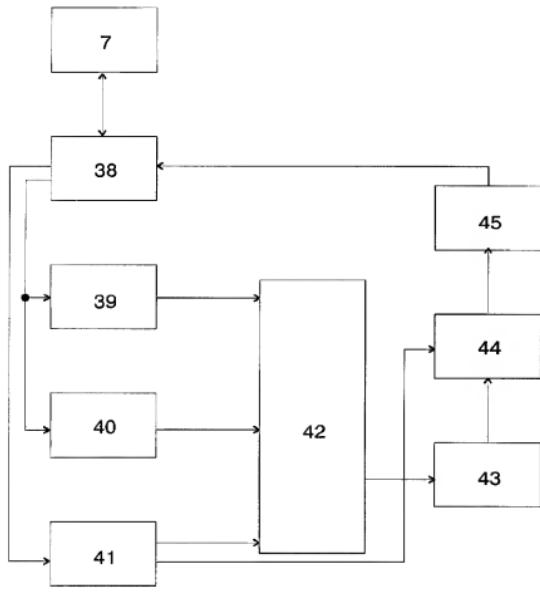
ФИГ. 5

РУ 2138076 С1



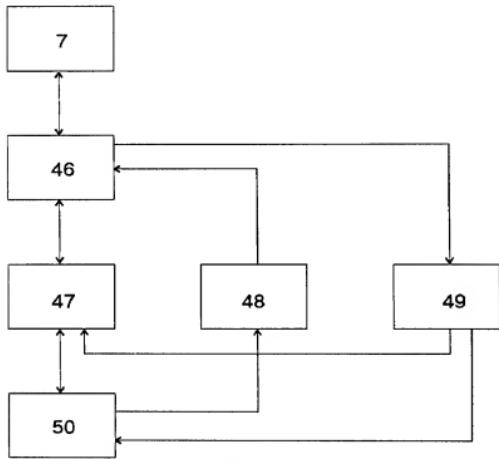
РУ 2138076 С1

ФИГ. 6



ФИГ. 7

РУ 2138076 С1



ФИГ. 8

РУ 2138076 С1